

ROMAN SCHNEIDER

Web 3.0 ante portas?

Integration von Social Web und Semantic Web

Das Medium Internet ist im Wandel, und mit ihm ändern sich seine Publikations- und Rezeptionsbedingungen. Welche Chancen bieten die momentan parallel diskutierten Zukunftsentwürfe von *Social Web* und *Semantic Web*? Zur Beantwortung dieser Frage beschäftigt sich der Beitrag mit den Grundlagen beider Modelle unter den Aspekten *Anwendungsbezug* und *Technologie*, beleuchtet darüber hinaus jedoch auch deren Unzulänglichkeiten sowie den Mehrwert einer mediengerechten Kombination. Am Beispiel des grammatischen Online-Informationssystems *grammis* wird eine Strategie zur integrativen Nutzung der jeweiligen Stärken skizziert.

1. Einleitung

Der Internet-Boom der letzten Jahre hat zu vielfältigen neuen sozialen Strukturen geführt, nicht zuletzt im www selbst. Das *Social Web* scheint, oft unter der griffigen Bezeichnung *Web 2.0*, inzwischen fest in der Medienrealität verankert. Was aber kommt danach? Welche alternativen oder weiterführenden Konzepte gibt es? Und wie können sie praktisch umgesetzt werden? Immer wieder diskutiert wird in diesem Zusammenhang eine vermeintliche Rivalität oder gar Unvereinbarkeit des Social Web mit den Paradigmen des *Semantic Web*, also mit der Idee der Speicherung von Internet-Inhalten in einer für Maschinen interpretierbaren und dadurch vielfältig weiterverarbeitbaren Form. Und ganz gleich, ob man diese Auseinandersetzung als spekulative Scheindebatte, als öffentlichkeitswirksame Marketing-Maßnahme oder aber als sachlich und fachlich existenzielle Grundsatzent-

scheidung auffasst, lässt sich eines kaum bestreiten: Dass die transdisziplinäre Diskussion in vielen Fällen auf unterschiedlichen Ebenen sowie unter Zuhilfenahme disparater Beurteilungskriterien geführt wird.

Der Leitgedanke des *Social Web* wird gemeinhin in etwa wie folgt formuliert: Jeder darf mitmachen, Privatsphäre und soziale Aktivitäten des Web-2.0-Surfers¹ verlagern sich zunehmend ins Internet, und die computervermittelte Interaktion zwischen den Teilnehmern ergänzt die bislang eher zentralisierte Erstellung von Inhalten. Praktische Ausprägungen dieser Ideen sind Weblogs, Podcasts, Wikis, Folksonomien oder Community-Portale – unkompliziert nutz- und erweiterbare Dienste also, die den Mediennutzer sowie dessen Bedürfnisse und Interessen verstärkt in den Mittelpunkt stellen.

Demgegenüber beruht die Vision vom *Semantic Web* auf der zunehmenden Akzeptanz und Verbreitung technologischer Standards, mit deren Hilfe sich Informationen im Internet formal so kodieren lassen, dass sie für Computerprogramme fassbar werden. Solche Standardisierungen betreffen beispielsweise Taxonomien, Ontologien und andere Metadatenmodelle für die Wissensrepräsentation bzw. die Charakterisierung von Web-Angeboten. Flankierend arbeitet eine globale Entwicklergemeinschaft an Protokollen und Modellen für applikationsübergreifende Interoperabilität.

Diese beiden Kurzcharakterisierungen illustrieren bereits die unterschiedlichen Blickwinkel sowie daraus resultierende Sprachprobleme, unter denen die Diskussion um den Stellenwert beider Ansätze leidet: Anwendungsorientiert hier (»Das Web als Plattform«) bzw. technologiebezogen dort (»Das Web als Datenbank«). Darüber hinaus scheint auch das Ziel nicht unumstritten: Geht es tatsächlich darum, ein in die Jahre gekommenes www abzulösen, oder nicht doch eher um die effektivere Nutzung bewährter Strukturen? Dieser komplexen Konstellation soll der vorliegende Beitrag Rechnung tragen, indem er die Perspektiven voneinander abgrenzt. In einem ersten Schritt werden Social Web und Semantic Web analysiert und beide Modelle auf eine mögliche Kombinierbarkeit hin untersucht. Konvergenz, Innovationspotenzial und Synergie-Effekte sollen dabei nicht nur theoretisch abgeleitet, sondern anhand eines Anwendungsbeispiels zumindest punktuell aufgezeigt werden.

1 Genauer: »Multi-Interest & User Generated Content«-Surfer, laut einer von Forsa erstellten repräsentativen Studie (vgl. SANDHÖFER 2007) eine von derzeit sieben Internet-Nutzergruppen.

2. Themen, Trends und Visionen im www

Es scheint ein Charakteristikum neuer Medientrends zu sein, ihr Innovationspotenzial aus den Schwächen etablierter Angebote abzuleiten. Neue Medien setzen sich demgemäß genau dann durch, wenn sie weitverbreiteten Kommunikationsbedürfnissen entgegenkommen, deren praktische Befriedigung aufgrund fehlender Basistechniken noch aussteht. Diese Erkenntnis beschränkt sich nicht auf den Bereich der digitalen Medien. Ob Gutenbergs Druckerpresse, die das Bedürfnis nach einer komfortablen Massenproduktion von Texten erfüllte, oder die technischen Grundlagen von Hörfunk oder Fernsehen: Stets fanden Innovationen gerade dann rege Verbreitung, wenn sie den Status quo für breite private und kommerzielle Interessensgruppen nachhaltig verbessern konnten.

Auch das World Wide Web (www) folgt bislang diesem Prinzip. 1989 von Tim Berners-Lee als Realisierung der Hypertext-Vision Ted Nelsons entworfen, revolutionierte es rasch die Kommunikationsrealität in einer immer stärker computerisierten Gesellschaft. Erste grafische Browser wie Mosaic oder Netscape erlaubten das bequeme Surfen; sie öffneten das Web für Benutzergruppen, denen die umständliche Bedienung früherer Internet-Applikationen noch unzumutbar erschienen wäre. Und abgesehen davon, dass das www bereits nach kurzer Zeit die Grenzen zu etablierten Medien verfließen ließ – Stichworte: Internet-Broadcasting oder Online-Zeitungen –, erlaubt es seit der sukzessiven Integration von Individualmedien wie E-Mail oder Telefonie erstmalig eine Verschmelzung von Individual- und Massenkommunikation. Während traditionelle Massenmedien den Austausch zwischen Autor und Rezipient nur innerhalb eines engen Rahmens erlauben, verspricht das www eine deutlich vereinfachte wechselseitige Kommunikation.

Ungefähr seit der Jahrtausendwende erscheinen nun Begriffe wie *Semantic Web* und *Social Web* bzw. *Web 2.0* auf der Agenda der fachübergreifenden Internetforschung. Beide Modelle verheißen, wie oben erwähnt mit unterschiedlichen Ansatzpunkten, eine qualitative Verbesserung des www. Auffallend ist, dass das »soziale« Web 2.0 seit seinem Ausruf (vgl. O'REILLY 2005) einen rasanten Aufstieg vom griffigen Trendbegriff hin zum praxisrelevanten Entwurf eines partizipativen Online-Mediums hinter sich hat. Die von Tim Berners-Lee u.a. 2001 formulierte Vision vom Semantic Web dagegen findet trotz vielfältiger und durchaus produktiver Aktivitäten bislang vergleichsweise selten den Weg aus dem akademi-

schen Bereich und den Forschungslabors großer Unternehmen wie IBM, HP und Oracle in die Medienrealität. Nach wie vor sind alltagstaugliche Praxisanwendungen Mangelware, und auch das öffentliche Interesse bleibt bestenfalls verhalten. Abbildung 1 illustriert diese auseinanderdriftende Entwicklung durch eine Visualisierung der weltweiten Suchanfragen bei der Internet-Suchmaschine *google*.

ABBILDUNG 1

Popularität der Begriffe ›Web 2.0‹ und ›Semantic Web‹ bei Google Trends (Stand September 2007)



2.1 Die Untersuchungsebenen

Um die Betrachtung von Social Web und Semantic Web auf eine tragfähige Basis zu stellen, empfiehlt sich vorab eine Darlegung der einzubeziehenden Untersuchungsebenen. Diese Auswahl muss naturgemäß selektiv bleiben und kann nicht sämtliche politischen, ökonomischen oder soziokulturellen Aspekte berücksichtigen. Derer gäbe es genug: Folgt nun das Social Web tatsächlich dem aus der Biologie bekannten Schwarmprinzip oder aber James Surowieckis sozialpsychologischer Theorie von der ›Weisheit der Vielen‹ (*wisdom of crowds*)? Befördern ›Social News‹ ohne professionelle Qualitätskontrolle den Niedergang des seriösen Journalismus oder ergänzen sich z.B. redaktionelle Hintergrundberichte und aktuelle Kundenempfehlungen zu einem ›Web Guided by Common Sense‹?² Und können in einem immer stärker vernetzten Informationsraum geistiges Eigentum und individuelle Privatsphäre noch geschützt werden oder ist das neue Web per se amoralisch?³

2 Vgl. den vielzitierten Online-Beitrag von John Markoff (2006) in der *New York Times*.

3 Vgl. z.B. Nicholas Carrs provokatives Essay (CARR 2005).

Der nachfolgende Vergleich von Social Web und Semantic Web stellt die in der informationswissenschaftlichen Fachdiskussion prominentesten Aspekte *Anwendungsbezug* und *Technologie* (vgl. hierzu z.B. ZIEGLER 2006) in den Mittelpunkt. Beide lassen sich gewinnbringend zur Charakterisierung von Webmodellen heranziehen. Allerdings gilt es, einen wesentlichen Umstand im Auge zu behalten: Das Social Web definiert sich *expressis verbis* aus dem Vorhandensein vernetzter bzw. vernetzungswilliger Teilnehmer, Dienste und Anwendungen. Es ist Realität, eben weil es eine beachtliche Zahl aktiver Nutzer aufweist, die das Etikett »sozial« erst begründen. Die Technologie stand anfänglich eher im Hintergrund; als elementar galt und gilt das Vorhandensein interaktiver und kollaborativer Elemente. Beinahe spiegelbildlich stellt sich die Entwicklung des Semantic Web dar. Der angestrebte Paradigmenwechsel bezieht sich dort eher auf die technische Infrastruktur, weniger auf eine – tatsächliche oder angenommene – Änderung des sozialen Verhaltens. Technologische Grundlagen, Spezifikationen und Standards stehen im Fokus der Arbeiten. Bis heute existieren nur wenige vermarktbarere Anwendungen, wohingegen der akademische und institutionelle Reifegrad stark fortgeschritten ist.

Zur technischen Untersuchungsebene gehören bei beiden Modellen die Themenbereiche *Wissensorganisation/Knowledge Retrieval* und mithin Fragen der terminologischen bzw. semantischen Interoperabilität. Für das Semantic Web – *nomen est omen* – leuchtet dies unmittelbar ein. Doch auch das Social Web beschäftigt sich bei näherer Betrachtung mit den Themen Wissensrepräsentation, -verarbeitung und -retrieval, beispielsweise in Gestalt verteilt erstellter Schlagworte (*tags*) und Begriffswolken (*tag clouds*).

2.2 *Das Social Web*

Zieht man die vielfältigen Partizipationsmöglichkeiten in offenen Gemeinschaften als zentrales Paradigma des Social Web heran, so wird rasch deutlich, dass sich die Ursprünge dieser Idee bis in die Anfangsjahre des www zurückverfolgen lassen. Ungeachtet dessen entfaltet sie erst seit wenigen Jahren eine Breitenwirkung, die sie für eine kritische Masse von Content-Providern attraktiv macht. Aktuelle Websites stehen in Sachen Benutzeradaptivität und Kollaboration für eine neue Quali-

tät; erst jetzt scheint das www tatsächlich interaktiv geworden zu sein. Die gesunkenen Zugangsbarrieren, primär technisches Vorwissen und Kosten für Breitbandanschlüsse, haben zu einer derart großen Anzahl engagierter Teilnehmer und sozialer Inhalte geführt, dass sowohl nicht-kommerzielle als auch profitorientierte Dienste Nutzen daraus ziehen. Das Ganze ist keineswegs nur Liebhaberei, sondern Wettbewerbsfaktor; es ist mehr als die Summe der Einzelteile und nicht länger nur eine Ansammlung disparater Informationseinheiten. Ob dies insgesamt das aus der Software-Entwicklung entlehnte Hantieren mit Versionsnummern begründet – Web 2.0 *suggestiert* ja eine signifikante technische und funktionale Innovation –, sei einmal dahingestellt.

Das Anwendungspotenzial ist auch deshalb enorm, weil sich Aspekte wie kontextabhängige Personalisierung und Benutzerpartizipation mittlerweile merklich ausgefeilter realisieren lassen, als es die bekannten Vorreiter wie *Yahoo* in den Neunzigerjahren vermitteln konnten. Das aktuelle Spektrum reicht von originären Social-Web-Projekten bis hin zur Einflechtung interaktiver Erweiterungsmöglichkeiten und individueller Profil-Auswertungen in »klassische« Websites. Insgesamt können folgende – nicht durchgehend disjunkte – Kategorien »sozialer« Anwendungen unterschieden werden:

Social Networking: Hierunter verstehen wir die momentan populärste Kategorie von Social-Web-Sites, deren Vertreter vormals wohl als C2C (Consumer-to-Consumer)-Portale oder schlicht Onlineforen firmiert hätten. Zentrales Element ist der Meinungsaustausch mit Hilfe von Weblogs, Podcasts, Wikis etc. Neben das kollektive Erstellen von Inhalten treten vermehrt auch gezieltes Eigenmarketing sowie die Abbildung sozialer Beziehungen und virtueller Lebensräume im Internet (*Blogger*, *MySpace*, *Friendster*, *Facebook*, *studiVZ*, *Xing/openBC* etc.).

Social Sharing: Multimediale Tauschbörsen fungieren als Treffpunkte für Anbieter und Interessenten von Videos, Tonaufnahmen, Fotos, Texten etc. Zentrale Dienste wie *Youtube*, *Flickr* und *Scribd* oder dessen deutschsprachiges Pendant *Doktus* erlauben neben Up- und Downloads auch die Verschlagwortung sowie individuelle Kommentierungen der eingestellten Inhalte.

Social News: Gelegentlich mit dem aufreizenden Etikett »Fünfte Gewalt« versehen, widmen sich unabhängige Newsportale wie *digg*, *yigg*, *netvibes* oder *wikio* der Aggregation, Verbreitung und Bewertung aktueller Nachrichten. Als Input dienen primär die etablierten journalistischen

Online-Angebote und -Nachschlagewerke, aber auch private Weblogs. Oft lassen sich die Startseiten durch das Ausfüllen von Präferenzformularen personalisieren.

Social Bookmarking: Das Sammeln und Veröffentlichen von Web-Lesezeichen ist ebenso alt wie das www. Neu hinzugekommen bei *del.icio.us* oder *Mister Wong* sind die Kombination individueller Kollektionen sowie das wechselseitige Verschlagworten. Oft sagt allein die Anzahl der Teilnehmer, die eine bestimmte Site in ihre Favoritenlisten aufgenommen haben, etwas über die Qualität – ganz sicher aber über die Akzeptanz – des betreffenden Angebots aus. Wie auch bei anderen Diensten im Social Web fließen die Ergebnisse des gemeinschaftlichen Indexierens in flach organisierte Wissensstrukturen ein.

Social Gaming: Der Spielebereich unterstreicht einen Multiplattform-Trend, der in Zukunft auch auf andere Kategorien übergreifen dürfte: Nicht allein der Webbrowser bietet Zutritt zur digitalen Unterhaltung im Internet, sondern zunehmend vermitteln mobile Geräte das simultane Treffen der Community-Mitglieder.

Social Commerce: Als Weiterentwicklung herkömmlicher B2C (Business-to-Consumer)-Modelle werden Geschäftsprozesse zwischen Verkäufer/Dienstleister und Kunden ins Web verlegt und um »soziale« Komponenten sowie Data-Mining-Prozeduren angereichert. Unternehmensportale beziehen den Verbraucher bzw. dessen aus Präferenzlisten, Navigationspfaden, Konsumverhalten oder anderen Online-Datenquellen zusammengestellten Charakteristika in die Werbungs- und Angebotsgestaltung ein. Nicht immer geschieht dies so transparent wie bei *Amazons* persönlichen Empfehlungen oder *Google AdSense*.

Die Grenzen zwischen diesen Kategorien sind fließend, Kombinationen (*mashups*) und Mehrfachverwendung (*content syndication*) eher die Regel als die Ausnahme. *BibSonomy* verbindet Social Sharing mit Social Bookmarking, *Wikipedia* weist Aspekte von Social Networking und Social News auf, und Dienste wie *Yahoo Pipes* erforschen innovative Integrationsoptionen für verteilte Webressourcen. Viele Anbieter fordern zwecks besserer Vernetzung – spricht: größerer Reichweite – ganz explizit zur Integration ihrer Funktionalitäten in andere Kontexte auf. Die *Flickr*- oder *Google Maps/OpenSocial-APIs* (*Application Programming Interfaces*) ermöglichen den externen Zugriff auf Portalinhalte und verbreitern damit die Basis für neue Dienste. Dabei ist allen genannten Beispielen eines gemein: Der Mehrwert gegenüber älteren Sites entsteht durch die »kol-

lektive Intelligenz«, d.h. durch die Aggregation individueller Beiträge, so klein sie im Einzelfall sein mögen, sowie durch das Mapping von Informationen. Durch den Übergang vom Push- zum Pull-Prinzip beteiligen sich die Nutzer aktiv an der Ausgestaltung der Angebote.

Aus technologischer Sicht greift das Social Web auf bereits eingeführte Einzeltechniken zurück, vereinzelt unter neuem Namen. Als markantes Beispiel hierfür sei Ajax (*Asynchronous JavaScript and XML*) genannt, ein Verfahren für den asynchronen Server-Client-Datentransfer und vormals unter der Bezeichnung *XMLHttpRequest* geläufig. Mit Hilfe dieser Techniken entstehen Browser-basierte Anwendungen, die sich in Design und Bedienungskomfort kaum von lokal installierter Software unterscheiden und Desktop-Funktionen zunehmend ins Web verlagern. Anerkennenswert ist die Präferenz für offene Standards. Dies bedeutet zwar nicht, dass sich sämtliche Web-2.0-Dienste der exakt selben Grundlagen bedienen, aber immerhin einen weitestgehenden Verzicht auf proprietäre Ansätze.

Die Vielfalt der Anwendungskontexte spiegelt sich in einer eindrucksvollen Format-Bandbreite wieder. Inhaltliche Strukturierungen werden zumeist unter Verwendung von XML-Sprachen⁴ durchgeführt – neben XHTML für Webseiten sind hier RSS/Atom für Abonnementdienste oder das Netzwerkprotokoll SOAP (*Simple Object Access Protocol*) für Web Services erwähnenswert. Unter der Bezeichnung »Mikroformate« (*micro formats* bzw. μF) verbreiten sich Ansätze (*hCard*, *hCalendar*, *hReview* etc.) zur Einbettung von Metainformation in XHTML-Instanzen, die ohne neuartige Element- oder Attributtypen auskommen. Dieses Vorgehen erlaubt z.B. die einfache Anreicherung von Webdokumenten um personen- oder ereignisspezifische Angaben, die gleichermaßen maschinenlesbar und für Menschen verständlich sind. Die Popularität von Mikroformaten schlägt sich in konkreten Anwendungen wie der Suchmaschine *Technorati Kitchen* nieder. Andere textbasierte Formate ohne XML-Hintergrund wie JSON (*JavaScript Object Notation*)⁵ haben dagegen eine vergleichsweise geringe Verbreitung.

Bereits seit der Einführung von Content-Management-Systemen (CMS) für die dynamische Generierung von Webseiten befindet sich der statisch programmierte Quellcode auf dem Rückzug. Weblog-Publikationssyste-

4 Damit verbunden sind natürlich die einschlägigen, mächtigen Technologien der weitverzweigten XML-Familie wie DOM, XSLT, XPath, XPointer, XQuery etc.

5 Ein interessantes aktuelles Anwendungsbeispiel für JSON ist die *Metaweb Query Language (MQL)* der Online-Datenbank *Freebase*.

me punkten zusätzlich mit Funktionalitäten zur Benachrichtigung über Trackbacks – Zitation von Blogbeiträgen an anderer Stelle – sowie einer feinkörnigen Rechte- und Rollenvergabe. Eine der bedeutsamsten Neuerungen dürfte die Schaffung gemeinschaftlicher Klassifikationsoptionen (*social* bzw. *collaborative tagging*) sein. Interessant ist hierbei insbesondere die Möglichkeit, die verteilt erstellten Schlagworte in sogenannten ›Folksonomien‹ (*folksonomies*, Wortschöpfung aus *folk* und *taxonomies*) zu sammeln und in Form von Begriffswolken – einer Art gewichteter Liste – zu visualisieren.

Die Tauglichkeit von ad-hoc-Folksonomien für die Recherche begründet sich eher durch die Menge der verfügbaren Deskriptoren sowie die breite Streuung der Mitwirkenden als durch die Qualität oder Effizienz der Klassifizierung. Da es weder eine zentrale Instanz für die Vergabe von Schlagworten noch ein kontrolliertes Vokabular gibt, etikettiert jeder Teilnehmer so, wie es ihm gerade in den Sinn kommt. Dies eröffnet die Chance, auch marginale Zusammenhänge zwischen einzelnen Etiketten bzw. Webinhalten aufzudecken und gleichzeitig dem ›Social Spamming‹ entgegenzuwirken. Orientierung bietet der Blick auf die Masse, d.h. die Hoffnung darauf, dass sich trotz aller Zersplitterungstendenzen die fachlich anerkannten und prägnanten Termini durchsetzen. Synonyme, orthografische Varianten, Abkürzungen oder Übersetzungen lassen sich in Folksonomien allerdings kaum zueinander in Beziehung setzen, Homonyme nicht ausreichend differenzieren. Erst ansatzweise bieten einzelne Dienste einen pragmatischen Ausweg aus dieser Misere. *Del.icio.us* beispielsweise erlaubt das manuelle Bündeln von Schlagworten; *Technorati* hat ein automatisiertes statistisches Verfahren zur Ermittlung ›verwandter‹ Deskriptoren eingeführt. Letzteres basiert auf der Auswertung hochfrequenter Kombinationen, also auf dem Umstand, dass z.B. eine signifikante Zahl von Teilnehmern eine Website wie *www.zeit.de* gleichermaßen mit den Etiketten ›Zeitung‹, ›Nachrichten‹ und ›Politik‹ belegt.

2.3 *Das Semantic Web*

Im Fokus der Arbeiten rund um das Semantic Web steht die computer-gestützte Verarbeitung von Webinhalten, insbesondere deren Analyse mit Hilfe universeller semantischer Strukturen und Verfahren. Es geht um die Verbesserung der Informationsqualität im weltweiten Netz durch

eine effizientere technische und semantische Interoperabilität: Wie formalisiert man komplexe Informationen auf beliebigen Webseiten so, dass nicht nur menschliche Nutzer unter Einsatz ihres Hintergrundwissens, sondern auch Computeralgorithmen produktiv darauf operieren können (Maschine-Maschine-Kommunikation)? Wie lassen sich heterogene Klassifikationsschemata und Terminologien integrieren oder portieren? Und an welchen Stellen unterstützt maschinelle Deduktion Verstehensprozesse und PC-basierte Alltagsaufgaben (Mensch-Maschine-Kommunikation bzw. maschinengestützte Interaktion zwischen Menschen)?


Der Mensch bleibt im Semantic Web also stets im Fokus der Aktivitäten. Im Gegensatz zu ambitionierteren Zielen der frühen KI-Forschung soll er nicht durch ›intelligente‹ Maschinen ersetzt, sondern lediglich von Fall zu Fall unterstützt werden. Dies übernehmen Softwarekomponenten, die durch eine Verknüpfung des explizit modellierten Welt- bzw. Domänenwissens mit zusätzlichen Einzelinformationen schlüssige Implikationen generieren. Die vielfältigen Anwendungsmöglichkeiten sind derzeit noch gar nicht erschöpfend analysiert. Prominente Demonstrationen und Produkte stammen beinahe durchweg aus dem Umfeld des Information Retrieval, und zwar aus gutem Grund: Konventionelle Internet-Suchmaschinen produzieren als Antwort auf Nutzeranfragen erfahrungsgemäß häufig absurde und scheinbar beziehungslose Ergebnislisten. Durch den Erfolg des Social Web mit all seinen Blogs, Podcasts und Wikis wird diese Problematik zukünftig sogar an Intensität zunehmen. Folksonomien und Begriffswolken bewirken hier wenig, denn sie unterstützen eher das Online-Flanieren als die präzise Recherche. Eine breitentaugliche, kommerziell verwertbare Retrievallösung dürfte folglich die Akzeptanz semantischer Technologien erheblich befördern.

Abbildung 2 illustriert eine entsprechende Anwendungssituation am Beispiel der Suchmaschine *CognitionSearch*.⁶ Unter Einsatz von linguistischen Analysealgorithmen und lexikalischem Wissen wird hier nicht einfach nach einzelnen Wortformen oder Begriffsklustern recherchiert, sondern zunächst die Bedeutung der einzelnen Anfragekomponenten bestimmt: *ebay* wird in der alternativen Schreibweise *e-Bay* automatisch als Auktionsplattform indentifiziert, *on-line* als Charakterisierung der

6 Einen alternativen Ansatz verfolgt die semantische Suchmaschine *hakia* unter Verwendung des NLP-Ansatzes *Ontological Semantics (OntoSem)*.

Zugangsweise ausgemacht usw. Der Anwender kann diese Annahmen optional nochmals überprüfen und erhält anschließend eine passgenaue Trefferliste.

ABBILDUNG 2
Semantische Technologien bei der Online-Recherche



HomePolitical Blogs? Help

For best Search results, use a plain English phrase or sentence with appropriate word capitalization.

Send Feedback

Search

The following word meanings were selected.
Use dropdown menus to change meanings.

e-Bay	1) n. an internet auction website
on-line	1) adj. adv. directly connected to or controlled by a computer
secure	1) adj. v. tr. safe; to fasten or close; secure the hatches / windows
shopping	1) n. visiting stores and looking at merchandise

Submit Changed Meanings

Advanced Query Edit

e-Bay plus on-line plus secure plus shopping: 8 files — displaying 8 results

Sort by RelevanceSort by Date

1. How to Report Online Fraud - Security Fix

Washington Post

I just got back from a very informative keynote discussion at ISPCON in Baltimore that I moderated with Howard Schmidt, the former Bush administration cyber-security [...]
http://blog.washingtonpost.com/securityfix/2005/05/how_to_report_online_fraud.html
► The following stems were matched exactly: e-Bay, on-line

2. eBay Invites Internet Regulation, Backs Online Gambling Ban

The Cato Institute

These companies are safe and reliable (FirePay is traded on the London Stock Exchange), but aren't subject to U.S. law, and so can be used for all sorts of goods [...]
http://www.cato.org/pub_display.php?pub_id=6429
► The following stems were matched exactly: e-Bay, on-line

Ausgangspunkt solcher semantisch fundierten Operationen zur Herstellung eines Realitäts- oder Objektbezugs ist das aus Sprachphilosophie und Linguistik bekannte semiotische Dreieck, d.h. die Unterscheidung zwischen *Ausdruck* (auch: *Zeichen*, *Wort*), *Konzept* (*Idee*, *Begriff*) und *Objekt* (*Bezeichnetes*, *Referent*). Um strukturierte Abfragen unter Nutzung eindeutiger Identifikatoren anstatt ambiger Stichworte durchzuführen, zieht semantische Software formalisierte Wissensbasen zu Rate. Diese speichern eingeführte Sachverhalte und Regeln und schaffen damit die Grundlage für eine automatisierte Überprüfung von Aussagen hinsichtlich ihres Wahrheitsgehalts, ihrer Redundanz oder Konsistenz. Wissensbasen können in Form monohierarchischer Taxonomien vorliegen oder als polymorphe Fachgebiets-Ontologien. Letztere werden aufgrund ihrer

Flexibilität und Mächtigkeit weithin favorisiert und weltweit in zumeist fachspezifisch organisierten Interessensgruppen aufgebaut.

Das World Wide Web Consortium (W3C) koordiniert die primär auf XML-affinen Technologien beruhende Infrastruktur des Semantic Web, insbesondere die Standards für Speicherung, Vernetzung und Abfrage. Formale Ontologie-Beschreibungssprachen sind der ISO-Standard *XML Topic Maps (XTM)*, *RDF Schema (RDFS)* sowie die *Web Ontology Language (OWL)*. Auf drei Sprachebenen (*OWL full*, *OWL DL*, *OWL lite*) lassen sich damit Konzepte bzw. Klassen, Eigenschaften, Instanzen und Beziehungen kodieren. Eine Verbindung von Ontologien mit Webdokumenten sowie die Cross-Referenzierung von Online-Inhalten erfolgt vorzugsweise in der Syntax des *Resource Description Frameworks (RDF)*. RDF arbeitet mit gerichteten Graphen und modelliert beliebige Aussagen als Tripel, bestehend aus Ressource (z.B. eine Webseite), Eigenschaft (Prädikat) und Objekt (Wert des Prädikats). Geschachtelte Aussagen werden durch die Kombination von Tripeln umgesetzt. Mehrere Arbeitsgruppen wie etwa die prominente *Dublin Core Metadata Initiative* setzen RDF bereits erfolgreich in der Praxis ein. Für die Extraktion von RDF-Inhalten sind verschiedene Abfrageverfahren in der Entwicklung; aussichtsreichster Kandidat für eine Standardisierung ist vermutlich das an die Datenbanksprache SQL angelehnte *SPARQL (SPARQL Protocol and RDF Query Language)*.

3. Integration – Ausweitung der Webzone?

Unsere Übersicht verdeutlicht, dass die Schwerpunkte Anwendungsbezug und Technikstandardisierung in Social Web und Semantic Web bislang in unterschiedlichem Maße verfolgt wurden. Aber gerade deshalb müssen beide Modelle nicht zwangsläufig als sich wechselseitig ausschließende Gegenpole betrachtet werden. Neben thematischen – Verbesserung der Such- und Präsentationsqualität – und technischen Überschneidungen – XML-basierte Formate und W3C-Empfehlungen – existieren eine Reihe sinnvoller Kombinationsmöglichkeiten. Infolgedessen stellt sich die Frage, warum das Gros der Web-Entscheider nicht ganz pragmatisch aus beiden Angebotstöpfen schöpfen sollte. Beispielsweise läge es nahe, Mikroformate zur Modellierungen simpler Strukturen und das mächtigere RDF für komplexere Zusammenhänge heranzuziehen. Neuerungen wie *GRDDL (Gleaning Resource Descriptions from Dialects of Languages)* – ein Mecha-

nismus zur Extraktion von RDF-Daten aus Webdokumenten – oder RDFa – erweiterte Annotationsattribute für Dokumentbeziehungen, Inhaltscharakterisierungen usw. – dürften diese Entwicklung befördern. Gleiches gilt für skos (*Simple Knowledge Organisation System*), einer Semantic Web-Spezifikation zur Modellierung einfacher Wissensstrukturen, die sich in der Praxis mühelos auf Folksonomien anwenden lässt.

Hinsichtlich einer echten Integration von Social Web und Semantic Web übernimmt die in beiden Ansätzen prominente inhaltliche Bewertung und Verknüpfung von Webinhalten zweifellos eine Brückenfunktion. Im Semantic Web findet Klassifikation gewissermaßen von oben herab statt, d.h., Experten modellieren sukzessive einzelne Weltausschnitte in Form domänenspezifischer Ontologien. Das Social Web dagegen setzt bei der inhaltlichen Klassifizierung auf einen konträren Bottom-up-Ansatz. Das ›Mitmach-Web‹ basiert auf der Möglichkeit jedes einzelnen Internet-Nutzers, Webinhalte eigenverantwortlich zu etikettieren. Niemand sitzt hinter den Kulissen und bewertet den Umstand, dass *Spiegel Online* von den einen mit den Schlagwörtern ›German Newspaper‹, von anderen mit ›Germany Online-News‹ oder ›Presse Nachrichten‹ charakterisiert wird (alle Beispiele von <http://del.icio.us>). Kreatives Chaos und lose Verknüpfungen also, gepaart mit der Hoffnung auf kollektive Intelligenz.

Wie sähe nun ein integrativer Ansatz in der praktischen Umsetzung aus? Das Institut für Deutsche Sprache (IDS) in Mannheim erforscht dies am Beispiel der Weiterentwicklung des datenbankgestützten grammatischen Informationssystems *grammis*. Gestartet Mitte der Neunzigerjahre als klassisches Webportal der ersten Generation, beinhaltet es neben einer hypertextuell organisierten ›Systematischen Grammatik‹ die Module ›Terminologisches Wörterbuch‹, ›Grammatisches Wörterbuch‹, ›Grammatische Bibliografie‹ sowie ›Grammatik in Fragen und Antworten‹. Ergänzend kamen später benutzeradaptive Funktionalitäten (vgl. SCHNEIDER 2004) sowie eine domänenspezifische Ontologie zur deutschen Grammatik hinzu. Bei der Kompilation der Ontologie unterstützten empirische Methoden den zeitaufwändigen intellektuellen Prozess der Wissenskodierung. Frequenz- und Mark-up-Analysen eines XML-kodierten Fachkorpus sowie mehrerer allgemeinsprachlicher Korpora lieferten den terminologischen Grundbestand; ein erprobter *weirdness*-Algorithmus berechnete Auffälligkeitswerte und ermittelte auf diese Weise domänenspezifische Konzeptkandidaten (vgl. SCHNEIDER 2006). Kollokationsanalysen unterschieden abschließend allgemeinere von spe-

zielleren Konzepten⁷, die seither von Hand gepflegt und erweitert werden. Die *grammis*-Ontologie steht also gewissermaßen für den ›state of the art‹ im Semantic Web.

Darauf aufbauend wird derzeit experimentell geprüft, wie semantische Informationen aus den einschlägigen Community-Portalen und anderen freien Quellen erfasst und eingebunden werden können. Erfahrungsgemäß versammeln sich in nutzergenerierten Online-Schlagwortlisten eher umgangs- als fachsprachliche Begriffe sowie unterschiedlichste Schreibvarianten. Die plausible Hypothese lautet: Ontologie (Semantic Web) plus Folksonomie (Social Web) mehrt das explizit verknüpfte elektronische Wissen und verbessert damit letztlich die Kommunikation zwischen Online-System und menschlichem Anwender. Insbesondere gestattet die Evaluation und Erweiterung von Ontologien um relevante Folksonomie-Begriffe⁸ eine benutzergerechtere Klassifikation und Filterung von Webinhalten.

ABBILDUNG 3

Beispiel einer Schlagwortliste und Begriffswolke bei *Mister Wong*



Konkret bedeutet das für *grammis* und andere innovative Informationsportale einen Ansporn zur statistischen Mustererkennung in Schlagwortlisten, Begriffswolken (vgl. Abb. 3), Wiki-Beiträgen oder struk-

7 Konzepte sind zusätzlich in ›bedeutungsähnlichen‹ Termsets und vermittels expliziter Relationstypen organisiert (vgl. SCHNEIDER 2007 sowie die Online-Version unter <http://www.ids-mannheim.de/grammis/>).

8 Erste Überlegungen zu einem kohärenten Modell einer ›TagOntology‹ finden sich z. B. in GRUBER 2005.

turierten RSS-Feeds. Methoden des Data Mining fördern signifikante Begriffskombinationen zutage und liefern erste Relevanz-Gewichtungen. Interessant wird dieser Ansatz nicht zuletzt durch den Umstand, dass eine Einbeziehung unterschiedlichster Folksonomien die aus der Semantic-Web-Forschung vertraute Problematik divergierender Modellierungs- und Terminologiesysteme aufgreift. Vermutlich lassen sich Konzepte und Beziehungstypen aus heterogenen Quellen zumeist durch eine intelligente Kombination algorithmisierter Ähnlichkeitsmessungen und manueller Nachbearbeitung aufeinander abbilden. Erfahrungen bezüglich der Integration von expliziten Wissensmodellen (*Ontology Mapping*)⁹ leisten hier wertvolle Dienste und treiben nebenbei die vielfach gewünschte Aggregation und Mehrfachnutzung sozialer Plattformen voran, so dass der Knowhow-Transfer zwischen Social Web und Semantic Web keine Einbahnstraße bleibt.

Abschließend sei auf einen nicht zu unterschätzenden positiven Nebeneffekt der Integration hingewiesen: Bislang stehen Folksonomien unter dem Generalverdacht, zu einer Simplifizierung komplexer Sachverhalte zu neigen, da sie überwiegend durch Laien aufgebaut und dürftig strukturiert sind. Auch Ontologien haben mit Akzeptanzfragen zu kämpfen – welches Modell bzw. welches Expertengremium ist für eine bestimmte Anwendungsdomäne tonangebend? Der Rückgriff auf die jeweiligen Stärken der beiden Modelle (*Net Credibility* vs. ausdrucksstarke Modellierung) sollte derartige Probleme entschärfen. Essenziell erscheint in diesem Zusammenhang die Einbeziehung individueller Rückmeldungen, also eines elementaren »sozialen« Charakteristikums. Gebräuchlich sind drei Varianten: a) explizite Voting- bzw. Rating-Systeme wie die Nutzerbewertung bei *ebay*, b) indirekte Mechanismen wie die auf den Bestellungen Dritter gestützten Kauftipps bei *amazon* oder die Weiterempfehlung von Artikeln auf Nachrichtenportalen sowie c) die als Spiel getarnte Profilermittlung. Unter der Prämisse einer möglichst breiten und repräsentativen Partizipation versprechen die beiden letztgenannten Rückmeldungstypen den größeren Erfolg (vgl. z. B. NIELSEN 2006). Dienste, die Innovationspotenziale aus Social Web und Semantic Web gleichermaßen aufgreifen wollen, werden zur Überprüfung ihrer Wissensrepräsentation und Retrievalergebnisse daher in Zukunft verstärkt auf die Analyse indirekter Rückmeldungen setzen.

9 Einen guten Überblick liefern z. B. STAAB/STUDER 2004.

4. Fazit

Der Abgleich technologischer und anwendungsbezogener Charakteristika von Social Web und Semantic Web bestätigt die Annahme, dass beide zumindest in Teilen ähnliche Ziele verfolgen, nämlich die informationelle Anreicherung von Webinhalten sowie die Verbesserung von Retrieval-ergebnissen. Es besteht also kein Interessenkonflikt zwischen den Modellen, und vieles spricht für eine Synthese der jeweiligen Vorteile. Technische Standards und semantische Kodierungsvarianten ergänzen sich eher, als dass sie sich gegenseitig im Wege stünden. Unterschiede existieren bezüglich der Realisierungsabfolge: Das Social Web ist eher benutzer- bzw. zielorientiert, während sich das Semantic Web als technologie- bzw. wegorientiert darstellt. Im Social Web ist die Mehrfachverwertung sowohl von Diensten als auch von Inhalten bereits weit verbreitet, während für das Semantic Web der Schwerpunkt auf der Wiederverwertung von Inhalten liegt.

Die Entscheidung, ob sich aus einer Kombination beider Modelle ein zukünftiges *Web 3.0* oder nur ein *Web 2.1 beta* entwickelt, bleibt weiterhin den Marketingexperten überlassen. Am Beispiel des Online-Informationssystems *grammis* konnte unabhängig davon angedeutet werden, welche Potenziale eine statistisch-texttechnologische Auswertung ›sozialer‹ Inhalte für wissensbasierte Systeme eröffnet. Zu berücksichtigen sind allerdings – neben Qualitäts- und Akzeptanzaspekten – stets auch juristische Fragen zum Urheberrecht, etwa bei der Analyse und Weiterverarbeitung von Inhalten aus Weblogs, Wikis oder Tauschbörsen.

Literatur

- BERNERS-LEE, T.; J. HENDLER; O. LASSILA: The Semantic Web. In: *Scientific American*, May 2001, S. 34-43
- CARR, N.: The amorality of Web 2.0. In: *Rough Type. Nicholas Carr's Blog*. 03. Oktober 2005, http://www.roughtype.com/archives/2005/10/the_amorality_o.php
- GRUBER, T.: *Folksonomy of Ontology: A Mash-up of Apples and Oranges*. First on-line conference on Metadata and Semantics Research (MTSRO5) 2005, <http://tomgruber.org/writing/mtsro5-ontology-of-folksonomy.htm>

- NIELSEN, J.: Participation Inequality: Encouraging More Users to Contribute. In: *Jakob Nielsen's Alertbox*, October 9, 2006, http://www.useit.com/alertbox/participation_inequality.html
- MARKOFF, J.: Entrepreneurs See a Web Guided by Common Sense. In: *The New York Times*, 12. November 2006, <http://www.nytimes.com>
- O'REILLY, T.: *What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*. 2005, <http://www.oreilly.de/artikel/web2o.html>
- SANDHÖFER, J.: *@facts extra – Online-Nutzertypen*. 2007, <http://www.Seven-OneInteractive.de>
- SCHNEIDER, R.: A Database-driven Ontology for German Grammar. In: REHM, G.; A. WITT; L. LEMNITZER (Hrsg.): *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Data Structures for Linguistic Resources and Applications*. Tübingen [Narr Verlag] 2007, S. 305-314
- SCHNEIDER, R.: Eine Ontologie für die Grammatik – Modellierung und Einsatzgebiete domänenspezifischer Wissensstrukturen. In: BUTT, M. (Hrsg.): *Proceedings of KONVENS 2006 (Konferenz zur Verarbeitung natürlicher Systeme)*. Konstanz [Universität Konstanz] 2006, S. 125-129
- SCHNEIDER, R.: *Benutzeradaptive Systeme im Internet*. Mannheim [Institut für deutsche Sprache – amades] 2004
- STAAB, S.; R. STUDER (Hrsg.): *Handbook on Ontologies*. Berlin, Heidelberg [Springer-Verlag] 2004
- ZIEGLER, C.: Smartes Chaos. Web 2.0 versus Semantic Web. In: *iX Magazin für professionelle Informationstechnik*, November 2006